Artificial Intelligence 172 (2008) 1800–1808

Contents lists available at ScienceDirect

Artificial Intelligence

www.elsevier.com/locate/artint

A note on the correctness of the causal ordering algorithm

Denver Dash [a,b,*,1], Marek J. Druzdzel [c,d]

[a] Intel Research Pittsburgh, 4720 Forbes Avenue, Pittsburgh, PA 15213, USA
[b] Department of Biomedical Informatics, School of Medicine, University of Pittsburgh, Pittsburgh, PA 15260, USA
[c] Decision Systems Laboratory, School of Information Sciences and Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA 15260, USA
[d] Faculty of Computer Science, Białystok Technical University, Wiejska 45A, 15-351 Białystok, Poland

ARTICLE INFO

Article history:
Received 28 October 2005
Received in revised form 16 June 2008
Accepted 19 June 2008
Available online 28 June 2008

Keywords:
Causality
Structural equation models

ABSTRACT

In this paper we examine in detail the algorithm of Simon [H.A. Simon, Causal ordering and identifiability, in: W.C. Hood, T.C. Koopmans (Eds.), Studies in Econometric Method. Cowles Commission for Research in Economics, Monograph No. 14, John Wiley & Sons, Inc., New York, 1953, pp. 49–74, Chapter III], called the *causal ordering algorithm* (COA), used for constructing the "causal ordering" of a system given a complete specification of the system in terms of a set of "structural" equations that govern the variables in the system. This algorithm constructs a graphical characterization of the model in a form that we call a *partial causal graph*. Simon argued in [H.A. Simon, Causal ordering and identifiability, in: W.C. Hood, T.C. Koopmans (Eds.), Studies in Econometric Method. Cowles Commission for Research in Economics, Monograph No. 14, John Wiley & Sons, Inc., New York, 1953, pp. 49–74, Chapter III] and subsequent papers that a graph so generated explicates causal structure among variables in the model. We formalize this claim further by proving that any causal model based on a one-to-one correspondence between equations and variables must be consistent with the COA.

1. Introduction

This note is concerned with a technique owing to Simon known as the *causal ordering algorithm* (COA). Given a self-contained system of simultaneous structural equations, COA will explicate asymmetries among variables in the system and produce a (possibly partial) matching between variables and equations. In a classic article [24], Simon showed that COA generates a directed graph which we call a *partial causal graph* (PCG). In [24] and in subsequent writings [11,12,25,26] Simon et al. argue that if a set of equations E is *self-contained* and composed of *causal mechanisms*, COA will produce causal graphs that are consistent with experts' "intuitive" causal orderings. We show in this note that the COA provides a summary of the necessary mappings from variables to equations. That is, *any* one-to-one mapping from variables to equations will be consistent with the COA. As a special case, when all clusters found by the COA contain only a single variable, then there exists only one mapping from equations to variables and only one (acyclic) directed causal graph, which is given by COA.

1.1. Preliminaries

For the purposes of this note, a causal model is defined as a set of equations together with a one-to-one mapping from equations to variables in the model (see Fig. 1). Matching a variable to an equation is an assertion that the other variables

* Corresponding author at: Intel Research Pittsburgh, 4720 Forbes Avenue, Pittsburgh, PA 15213, USA.
E-mail addresses: denver.h.dash@intel.com (D. Dash), marek@sis.pitt.edu (M.J. Druzdzel).

[1] This work was performed while Denver Dash was a graduate student at the Intelligent Systems Program, University of Pittsburgh.

0004-3702/$ – see front matter
doi:10.1016/j.artint.2008.06.005

$$
\begin{aligned}
&e_1\colon w = f_1(z, v, x) && (\text{maps to } w)\\
&e_2\colon x = x_0 && (\text{maps to } x)\\
&e_3\colon z = f_2(x, y) && (\text{maps to } z)\\
&e_4\colon y = f_3(z) && (\text{maps to } y)\\
&e_5\colon v = v_0 && (\text{maps to } v)
\end{aligned}
$$

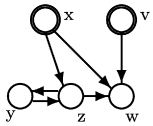


Fig. 1. A causal model is specified by a set of equations and a one-to-one correspondence between equations and variables and defines a directed graph over the variables.

present in that equation are causal parents of the matched variable. Such a specification of a system defines a directed graph (DiG) which is interpreted as a causal graph. This representation of causality has its roots in structural equation models in the econometrics literature [8,10,24,30,36,37] and has been developed further within AI over the past decade [6,21,29]).

It is frequently the case that we know the equations that govern a given system but we are unsure about the correct mapping from variables to equations. For example, any physical system typically has an associated set of physics equations that govern the processes in the system. Many socio-economic models have "laws" that are represented by equations that must be satisfied for a given set of assumptions. Given such a set of equations, in order to produce a causal model (as defined in this note), one must be able to generate a one-to-one correspondence between variables in the system and equations. To do this in general requires detailed background knowledge about causal interactions in the system. In practice for even modest systems it can become an intractable task by hand. It would thus be desirable to possess an automated method by which the matching between equations and variables can be generated. Such a method would be especially valuable for very large models possessing hundreds or thousands of variables. A central practical problem with such an automated method is to ensure that the mapping generated has causal meaning.

Our contributions here are to formally define a Partial Causal Graph (PCG) and define a notion of consistency between a DiG and a PCG. We then show that the mapping of equations to variables produced by COA will be consistent with any other mapping, and thus the PCG generated by COA will be consistent with any DiG that is consistent with E. By "consistent" we mean in essence that any arc present in the PCG must be present in the DiG, and any arc in the DiG must not be ruled out by the PCG. Our proof requires neither linear equations nor equations which can be solved for unique values for the variables.

We feel that this work is significant because it serves to validate decades of research which has shown COA to be a powerful tool for operating on causal models. One of the primary uses for causal graphs in general is to support the ability to reason about the effects of manipulation on a real-world system and predict the resulting probability distribution. The *Do* operator of Pearl [21] is a well-known case of an operator for modeling manipulation of a variable when such a manipulation breaks the connection between the variable and its parents. However, the COA has served as a generating function for all sorts of operations on causal models. For example, COA can be used to model the restructuring that occurs in a dynamic causal system when it passes through equilibrium [2,3,11]. Yet another operation might be the replacement of some components with others that depend on qualitatively different factors, such as replacing a spring with a compressible gas piston. COA is capable of modeling manipulation when reversible mechanisms are present in the model [7]. This technique was used in a model for strategic business planning by the administration at Carnegie Mellon University [23]. Given a library of fundamental laws describing an arbitrary system, COA also provides a method to automate the process of model building by constructing causal graphs on the fly, depending on which devices are added to the system [16]. The validity of using COA for these purposes, however, rests on the existence of a proof of the correctness of COA. Thus, the key significance of this paper is that it converts an entire thread of research from a set of useful heuristics to provably correct techniques.

1.2. Previous work

Much work on causality has been performed in the past decades in statistics and artificial intelligence. This work has been concerned with representation (e.g., [13,14,19,33,34]), inference (e.g., [15,19]), causal reasoning (e.g., [20,28]), learning from data (e.g., [1,18,27,28]), among other topics. Most of this work has dealt with causal models that are very similar to the type constructed with the COA; however, in their formulations, a causal model is assumed as a given or it is derived from data, and the process of converting a set of equations to a causal model is not considered.

Nayak [17] comes the closest to addressing the question that we pose here. He shows that all mappings between structural equations and variables produce the same set of ancestor-descendant pairs. Similarly, we will show that all mappings possess common features, but these common features will be in terms of direct causal connections rather than indirect ancestral relations, and we provide the proof that COA provides a condensed representation of those necessary direct connections. Dash [2,3] shows that the causal interpretation of equilibrium systems is not straightforward due to the fact that underlying dynamics can lead to equilibrium independence graphs that are not causal. He terms this reason for non-causality "violation of Equilibration–Manipulation commutability." We emphasize that the work presented here does not imply that models retrieved by COA are assured to obey the Equilibration–Manipulation commutability property. There exist many other concepts of causality that do not involve a mapping from equations to variables. Granger causality [9] uses

correlations across time to identify causal relations. The work by de Kleer and Brown [4] and that of Williams [35] address the problem of determining causality from a set of constraints by propagating disturbances on variables in the model. There is a debate as to whether the formalisms of de Kleer and Brown and Williams are consistent with the concept of causality used in this note [5,12]. We make no claim that the proofs presented in this note apply to Granger causality or causal representations used by other work (cf. [22,31,32]), unless those other formalisms can be expressed as a one-to-one mapping between equations and variables.

Iwasaki and Simon [11] extend the COA to construct dynamic causal models given a set of differential equations specifying the system. The addition of differential equations causes their modified version of the COA to add additional "integration arcs" to the graph produced by the original COA. However, the graph obtained by their algorithm prior to the addition of these special arcs is identical to the graph that would be obtained by the COA on the given system of equations; therefore, the results that we present here apply equally well to the static portion of their dynamic causal ordering. Also, our result directly applies to their demonstration of how equilibration affects the causal ordering of a system, again if one omits the integration arcs from the analysis. We conjecture that our proofs can also be applied to the integration arcs if one considers them to denote causation across time. To prove this, one needs to show how new variables can be introduced to the system denoting future versions of existing variables, and how a differential equation can be expanded to cover these new variables.

1.3. Notation

We will use the following notation throughout the remainder of the note: When it does not conflict with other notation, we will denote elements of sets by lower case letters and sets by upper case letters. If $G = \langle V, A \rangle$ is a directed graph where V is a set of vertices and A is a set of directed arcs, we will use $\mathrm{Pa}(v)_G$ and $\mathrm{Ch}(v)_G$ to denote the parents and children, respectively, in G, for some $v \in V$. We will use $\mathrm{Anc}(v)_G$ and $\mathrm{Des}(v)_G$ to denote the ancestors and descendants of v in G. In all cases, we may drop the subscript G if the graph is implied by the context. If e is an equation then we use $\mathrm{Params}(e)$ to denote the set of variables constrained by e. If E is a set of equations $e_1, e_2, \ldots, e_n$, we use $\mathrm{Params}(E)$ to represent $\bigcup_{e_i \in E} \mathrm{Params}(e_i)$.

A partition X_p of a set X is a set of disjoint sets $X_p = \{X_1, X_2, \ldots, X_n\}$ such that $\bigcup_{i=1}^{n} X_i = X$; we call the disjoint sets of a partition the *clusters*. For example $\{\{1, 3, 5\}, \{2, 4, 6\}, \{7\}\}$ is a partition of {1,2,3,4,5,6,7}, and {1,3,5} is a cluster. If X_p is a partition of a set X and $x \in X$, then we use the notation $\mathrm{Clust}(x)_{X_p}$ to denote the cluster in which x lies in X_p. We will drop the subscript X_p from $\mathrm{Clust}(x)_{X_p}$ if the partition is clear by the context.

1.4. Causal modeling and COA: Examples

We are considering causal models in the form of *structural equation models* (SEMs), whereby a system, summarized by a set of feature variables V, is specified by a set of equations E which determines a complete solution set for V, and each variable $v \in V$ is associated with a single unique equation $e \in E$. Such a specification of a causal system defines a directed graph over the variables by defining the parent set of v to be the remaining parameters of e which are also in V. An example of such a model is shown in Fig. 1.

Some variables in V may be determined outside of the system, meaning that, although they are tied to the system in some way (because they appear in some $e \in E$), their values are determined either by external, uncontrollable influences (e.g., the weather) or they are under direct experimental control (e.g., the temperature of a gas). These variables are called *exogenous*. Variables x and v in Fig. 1 are exogenous. A variable x is exogenous if and only if there exists an equation that can be written in the form $f(x) = 0$, where f is some function. For example the equation $x = x_0$ can be rewritten as $x - x_0 = 0$. All variables in the model that are not exogenous are called *endogenous*, meaning that their values are determined within the system.

Typically it is assumed that if an equation e is associated with a variable v, then v can be written as a function of the remaining parameters of e: $v = f(\mathrm{Pa}(v))$. However, this restriction is not necessary in general. The alternative is that the equation e will form an implicit equation for v given values for $\mathrm{Pa}(v)$, and will thus constrain the outcomes of v without necessarily determining a unique value.

The causal ordering algorithm (COA) [24] is an automated method that, given a set of equations E over variables V, produces a matching between equations and variables. In the process, it constructs a *partial causal graph*. COA generates a matching by recursively associating to each variable v the smallest subset of equations that can be solved in order to determine the value for v given the variables that have already been determined. An example is shown in Fig. 2: initially, the values of x and v can be determined by solving equations e_2 and e_5, each being an equation in one variable. Thus, x is associated with e_2 and v with e_5. Now, given values for x and v, no remaining variables can be determined within a single equation (because, for example, there is no equation like $f(x, y) = 0$). However, the equations e_3 and e_4 taken together as a subsystem can be simultaneously solved for all variables appearing in e_3 and e_4 together. Simon called $\{e_3, e_4\}$ a *minimal self-contained subset* of equations, because there is no smaller subset of equations that can be solved for its parameters. For a fixed value of x, this subset contains exactly two variables, y and z, which can both be simultaneously solved for in terms of x. Thus the set of variables $\{y, z\}$ is associated with the set of equations $\{e_3, e_4\}$. The variables y and z are said to be *strongly coupled* because neither one could be determined before the other, and they are collectively represented by a single

$$
\begin{aligned}
&e_1\colon\ g_1(x,v,z,w)=0 && \text{maps to } w\\
&e_2\colon\ g_2(x)=0 && \text{maps to } x\\
&e_3\colon\ g_2(x,y,z)=0 \\
&e_4\colon\ g_3(y,z)=0 && \text{maps to } \{y,z\}\\
&e_5\colon\ g_4(v)=0 && \text{maps to } v
\end{aligned}
$$

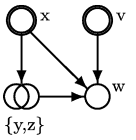
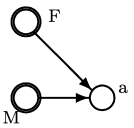


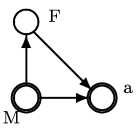
Fig. 2. The causal ordering algorithm takes an unordered set of equations as input and outputs a partial matching between variables and equations, which defines a *directed cluster graph*.

$$
\begin{aligned}
F &= 10\,N\\
M &= 5\,kg\\
a &= F/M
\end{aligned}
$$



(a)

$$
\begin{aligned}
F &= (1m/s^2)M + 15N\\
M &= 5\,kg\\
a &= F/M
\end{aligned}
$$



(b)

Fig. 3. Two equation systems and the corresponding causal graphs obtained by the COA. Although both sets of equations are algebraically equivalent, the structures obtained are different. *N*, *kg m* and *s* stand for the units Newtons, kilograms, meters and seconds, respectively.

vertex in the PCG, as in Fig. 2. Since x appears in e_3, it is a parent of the entire subset $\{y, z\}$ in the graph. Finally, given x, v, and z, the value of w is determined in e_1, so w is associated with e_1, and x, v and z are parents of w in the graph.

All graphs produced by COA will be acyclic, but they may possess strongly coupled vertices as in Fig. 2, which, if resolved by manually associating all strongly-coupled variables with individual equations, could create cycles in the resulting directed graph (e.g., see Fig. 1). The models and graphs produced by COA are thus not typically what is referred to as "causal models" and "causal graphs" in the literature on causality (Simon's work excluded). On the contrary, a causal model typically requires an association between single variables and single equations which in turn defines a DiG, where each vertex of the graph corresponds to a single variable rather than a set of variables. Also note that, despite the fact that a PCG is a directed graph involving clusters of variables, it is not a typical clique graph. Although incoming arcs are associated with the entire cluster, the outgoing arcs are specified as variable-to-cluster arcs, not cluster-to-cluster.

The ability of a causal model to represent the structure of a system of equations depends critically on the fact that each equation is *structural*. A structural equation differs from a normal algebraic equation in that it is not considered identical to all other equations logically entailed by it. Likewise, a system of structural equations is not considered identical to all systems of equations entailed by it.

This restriction prohibits a SEM from undergoing arbitrary algebraic manipulations, because this will in general alter the causal ordering of the system by creating different structural equations. Thus, in SEMs, the algebraic arrangement of equations in the system determines the causal structure. Consider, for example, the causal model for a mass undergoing acceleration due to Newton's 2nd law in Fig. 3. In Fig. 3(a), the structural equation model and corresponding PCG are shown for the case when the force F is set to a value of 10 Newtons and the mass M is set to a value of 5 kg. On the other hand, the model in Fig. 3(b) contains a set of equations that is algebraically equivalent to the set in Fig. 3(a); however, the only way to assign variables to the equations in Fig. 3(b) results in a totally different structure. As not all of the equations in this system are structural (the first equation in Fig. 3(b) is an algebraic combination of the first two equations in Fig. 3(a)), the structure cannot be given a causal interpretation.

If one is given the set of equations in Fig. 3(a), it is easy to see that the only way to map variables to equations results in the causal graph shown there, and it is intuitive that this is the correct causal ordering if the exogenous variables are interpreted as being *set* to their particular values. At it's core, COA is as simple as being able to match up equations to variables that have not been determined by the system yet. In this example it is easy to see how to do that, but in large, complex systems, this task can quickly get overwhelming without an automated procedure.

Our ultimate goal in this note is to compare the PCGs produced by COA with graphs produced by an arbitrary matching of variables to equations (e.g., comparing the PCG of Fig. 2 with the DiG of Fig. 1). We will show that any fully-specified matching must be consistent with the partial matching produced by COA. However, a comparison of the two methods is not trivial if only because they produce different graphical structures. We therefore must go through some pains just to bring the two methods onto the same level so that they can be compared. In Section 2, we present formal definitions that will be used to facilitate a comparison of the two approaches.

2. The causal ordering algorithm

In this section, we will present COA formally, which is necessary to understand our proof. We refer the reader to Simon [24] for a less concise treatment. We first define a set of restrictions that must apply to a set of equations that comprise a model:

Definition 1 *(self-contained structure).* A set of equations E is *self-contained* iff $|E| = |\text{Params}(E)|$, and for every subset $E' \subset E$, $|E'| \leqslant |\text{Params}(E')|$.

Note that by assuming that a structure is self-contained we (and COA) conveniently exclude problematic systems of equations (e.g., those for which there are more equations than variables or visa-versa). Also, this definition only considers the qualitative structure present in equations. Some self-contained structures make no sense as a causal system. For example, the following are self-contained structures:

$$X = 2Y$$

$$10X = 20Y$$

Obviously such a system of equations can never completely represent a real causal system. Apparently enough background knowledge must exists to prevent the inclusion of equations that are similarly contradictory or non-independent, but our theorems apply regardless as long as the system is self-contained.

Definition 2 *(minimal self-contained set).* If S is a self-contained structure, then S is a *minimal self-contained set* iff there does not exist a subset $S' \subset S$ such that S' is also self-contained.

Note that an equation for a single variable such as $x = x_0$ will form a (minimal) self-contained subset. If E is a self-contained structure, then E defines a unique set $\mathcal{S}$ of solutions over the variables in $\text{Params}(E)$. If all variables have unique solutions, then $|\mathcal{S}| = 1$. Our definition of a self-contained structure varies slightly from that in [24] because our formulation generalizes to non-linear equations with non-unique values, similarly to the work of Iwasaki and Simon. COA, as we define it, does not need to actually solve equations for values. It is only concerned with variable dependencies. If a non-linear equation leads to an implicit solution to a variable, then the causal system can be viewed as simply constraining that variable based on causal parents rather than absolutely determining its value.

COA recursively constructs *derived subsets* by finding minimal self-contained sets and reducing the set of equations by substituting out those variables found so far:

Definition 3 *(derived subset).* Let E be a self-contained structure and let $\mathcal{E}_{sc}$ be the set of all minimal self-contained subsets of E. Let $E_{sc} = \bigcup_{E_i \in \mathcal{E}_{sc}} E_i$, and let $E_{\leftarrow}$ denote the set of equations that are obtained when values of $\text{Params}(E_{sc})$ consistent with E_{sc} are substituted into $E \setminus E_{sc}$. $E_{\leftarrow}$ is called the *derived subset* of E.

We will use the notation that $E^{(1)} \equiv E_{\leftarrow}$, $E^{(2)} \equiv E^{(1)}_{\leftarrow}$, etc., where $E^{(i)}$ is called the *derived subset of ith order*. If E is a set of equations with derived subset $E^{(i)}$, and if $E' \subseteq E^{(i)}$ is some subset of $E^{(i)}$, then we use $\hat{E}'$ to denote the subset of E corresponding to the equations remaining in E', i.e., the subset of original equations with no values substituted.

Definition 4 *(well-defined self-contained structure).* If E is a self-contained structure, then E is *well-defined* iff the derived subset $E^{(i)}$ is self-contained, for all i.

For our purposes, we assume that all self-contained structures are well-defined.

In Fig. 2, COA constructed a mapping between sets of variables and sets of equations. The mapping could be written as a list of associations as follows: $\phi = \{\langle\{w\}, \{e_1\}\rangle, \langle\{x\}, \{e_2\}\rangle, \langle\{y, z\}, \{e_3, e_4\}\rangle, \langle\{v\}, \{e_5\}\rangle\}$. We say that the partitions $\{\{w\}, \{x\}, \{y, z\}, \{v\}\}$ and $\{\{e_1\}, \{e_2\}, \{e_3, e_4\}, \{e_5\}\}$ are *commensurate* and ϕ defines a *partial causal mapping*:

Definition 5 *(commensurate partitions).* Let A and B be two sets such that $|A| = |B|$. A partition P_A over A is *commensurate* with a partition P_B over B iff there exists an onto mapping $\phi : P_A \to P_B$ such that for each set $S^{(i)}_A \in P_A$, $|S^{(i)}_A| = |\phi(S^{(i)}_A)|$.

Commensurate partitions are obviously one-to-one.

Definition 6 *(partial causal mapping).* If E is a self-contained set of equations with $V = \text{Params}(E)$, then a *partial causal mapping* Φ of E is a triple $\langle V_p, E_p, \phi\rangle$, where V_p is a partition of V, E_p is a partition of E, and ϕ is a bijection, $\phi : V_p \to E_p$, such that the following is true:

(1) $\{V_p\}$ is commensurate with $\{E_p\}$, and
(2) For all sets $V^{(i)}_p \in V_p$ we can match up each variable in $V^{(i)}_p$ with a unique equation in $\phi(V^{(i)}_p)$. That is, there exists a bijection $\phi^{(i)} : V^{(i)}_p \to \phi(V^{(i)}_p)$ such that $x \in \text{Params}(\phi^{(i)}(x))$ for all $x \in V^{(i)}_p$.

A partial causal mapping $\Phi = \langle V_p, E_p, \phi \rangle$ can also be written as a list of ordered pairs or *associations*: $\Phi = \{\langle V_p^{(1)}, E_p^{(1)} \rangle, \langle V_p^{(2)}, E_p^{(2)} \rangle, \ldots, \langle V_p^{(n)}, E_p^{(n)} \rangle\}$, where $V_p^{(i)} \in V_p$ and $E_p^{(i)} \in E_p$ are sets for all i, and n is the number of clusters in V_p. We will use these two representations of a partial causal mapping interchangeably.

COA takes a well-defined self-contained structure and converts it to a partial causal mapping:

Definition 7 *(Causal Ordering Algorithm).* Given a well-defined self-contained structure $E^{(0)}$, the *Causal Ordering Algorithm* produces a partial causal mapping $\Phi = \langle V_p, E_p, \phi \rangle$ over $E^{(0)}$ through the following procedure:

(1) Define $\mathcal{E}_{sc}^{(i)}$ to be the set of all minimal self-contained subsets of $E^{(i)}$.
(2) For each set $E \in \mathcal{E}_{sc}^{(i)}$ add the association $\langle \text{Params}(E), \hat{E} \rangle$ to Φ.
(3) Let $E^{(i+1)} \equiv E_{\hookleftarrow}^{(i)}$ and recurse this procedure until $E^{(i+1)} = \emptyset$.

We use COA(E) to denote the partial causal mapping generated by applying COA to the equation set E.

The partial causal mapping defined by COA in Fig. 2 also defined the directed graph shown in the figure. We call this type of graph a *partial causal graph*:

Definition 8 *(partial causal graph).* A partial causal graph is an ordered pair $\langle V_p, A_p \rangle$, where the set of vertices V_p is a partition of V and A_p is a set of directed arcs $v \to V_p^{(i)}$ where $v \in V$ is a variable, $V_p^{(i)} \in V_p$ is a cluster, and $V_p^{(i)} \neq \text{Clust}(v)$.

The PCG associated with a partial causal mapping $\Phi = \langle V_p, E_p, \phi \rangle$ can be constructed as follows: For each association $\langle V_p^{(i)}, E_p^{(i)} \rangle \in \phi$ and for each $e \in E_p^{(i)}$ and each $v \in \text{Params}(e) \setminus V_p^{(i)}$, direct an edge from v to $V_p^{(i)}$.

A causal graph can be viewed as a special case of a PCG, where all associations of the partial causal mapping are *elementary*:

Definition 9 *(elementary association).* We will call an association $\langle V_p^{(i)}, E_p^{(i)} \rangle$ an *elementary association* if $|V_p^{(i)}| = |E_p^{(i)}| = 1$. If $\langle V_p^{(i)}, E_p^{(i)} \rangle$ is an elementary association where $V_p = \{v\}$ and $E_p = \{e\}$, for clarity of notation we will often write this association as $\langle v, e \rangle$ rather than as $\langle \{v\}, \{e\} \rangle$.

Definition 10 *(total causal mapping).* If E is a set of equations with $V \equiv \text{Params}(E)$, then a *total causal mapping* over E is a bijection $\phi : V \to E$.

Nayak [17] proves that a set of independent equations is self-contained iff it possesses a total causal mapping. In the same way that a partial causal mapping can be used to construct a PCG, a total causal mapping defines a directed (possibly cyclic) graph (DiG):

Definition 11 *(causal model).* A *structural equation model* S is a triple $S = \langle E, V, \phi \rangle$, where E is a self-contained structure over parameters V, and $\phi : V \to E$ is a total causal mapping.

We use the terms *causal model* and *structural equation model* interchangeably. We denote the DiG that corresponds to a total causal mapping ϕ_t as DiG(ϕ_t) and the PCG that corresponds to a partial causal mapping Φ_p as PCG(Φ_p)

3. Correctness of COA

Comparison between the causal ordering given by COA and that given by an arbitrary expert is complicated by the fact that both procedures produce different types of directed graphs. We, therefore, must define precisely what we mean when we say that a DiG is consistent with a PCG.

Definition 12 *(DiG/PCG consistency).* If V is a set of variables, $G_p = \langle V_p, A_p \rangle$ is a PCG over V, and $G = \langle V, A \rangle$ is a DiG over V, then G is *consistent* with G_p if and only if the following are true:

(1) If an edge $v_1 \to V_p$ exists in A_p then there exists a $v_2 \in V_p$ such that the edge $v_1 \to v_2$ exists in A.
(2) If an edge $v_1 \to v_2$ exists in A, then either $v_1 \to \text{Clust}(v_2)$ exists in A_p or $\text{Clust}(v_1) = \text{Clust}(v_2)$.

Condition 12.1 says that all arcs present in G_p must be represented in G, and Condition 12.2 says that the only additional arcs that are allowed must be between variables that were "strongly coupled" in G_p. The DiG of Fig. 1 is consistent with the PCG of Fig. 2.

Definition 13 *(partial/total mapping consistency).* Let Φ_p be a partial causal mapping over a self-contained structure E with $V = \text{Params}(E)$, and let ϕ_t be a total causal mapping over E. ϕ_t is *consistent* with Φ_p iff the following hold:

(1) For each association $\langle V_p^{(i)}, \Phi_p(V_p^{(i)})\rangle \in \Phi_p$, there exists for each $v \in V_p^{(i)}$ an elementary association $\langle v, e\rangle \in \phi_t$, where $e \in \Phi_p(V_p^{(i)})$.
(2) An elementary association $\langle v, e\rangle$ exists in ϕ_t but not in Φ_p only if the non-elementary association $\langle \text{Clust}(v), \text{Clust}(e)\rangle$ exists in Φ_p.

The following lemma shows that mapping consistency implies DiG/PCG consistency:

Lemma 1. *Let E denote a well-defined self-contained structure. If a total causal mapping ϕ_t over E is consistent with a partial causal mapping $\Phi_p = \langle V_p, E_p, \phi_p\rangle$ over E, then* $\text{DiG}(\phi_t)$ *is consistent with* $\text{PCG}(\Phi_p)$.

Proof. Let $G_p = \langle V_p, A_p\rangle$ denote $\text{PCG}(\Phi_p)$ and let $G_t = \langle V, A\rangle$ denote $\text{DiG}(\phi_t)$. Assume that conditions 13.1 and 13.2 are true.

Satisfaction of condition 12.1:

Assume an edge $v_1 \to V_p^{(2)}$ exists in A_p. Let $\langle V_p^{(2)}, \phi_p(V_p^{(2)})\rangle$ be the association corresponding to $V_p^{(2)}$ in ϕ_p. By condition 13.1, there exists in ϕ_t an elementary association of the form $\langle v_2, e_1\rangle$ where $v_2 \in V_p^{(2)}$ and $e_1 \in \phi_p(V_p^{(2)})$. Therefore in $\text{DiG}(\phi_t)$ there exists an edge from all $v_1^i \in \text{Params}(e_1) \setminus V_p^{(2)}$ to some $v_2 \in V_p^{(2)}$. Finally, since $v_1 \to V_p^{(2)}$ it must be the case that $v_1 \in \text{Params}(e_1) \setminus V_p^{(2)}$.

Satisfaction of condition 12.2:

Assume an edge $v_1 \to v_2$ exists in A, then the elementary association $\langle v_2, e_1\rangle$ must exist in ϕ_t such that $v_1 \in \text{Params}(e_1)$. Then by condition 13.2, either $\langle\{v_2\}, \{e_1\}\rangle \in \phi_p$ or $\langle \text{Clust}(v_2), \text{Clust}(e_1)\rangle \in \phi_p$. Either way the association $\langle \text{Clust}(v_2), \text{Clust}(e_1)\rangle \in \phi_p$. Therefore in $\text{PCG}(\Phi_p)$ an arc will be directed from all $v \in \text{Params}(\text{Clust}(e_1)) \setminus \text{Clust}(v_2)$ to $\text{Clust}(v_2)$. Therefore since $v_1 \in \text{Params}(\text{Clust}(e_1))$, either there will exist an edge $v_1 \to \text{Clust}(v_2)$ or $v_1 \in \text{Clust}(v_2)$. □

Using this result, Theorem 1 shows that a DiG G_t generated by any total causal mapping ϕ_t over a set of equations E is consistent with the PCG G_p generated by applying COA to E:

Theorem 1. *Let E be a well-defined self-contained structure, let ϕ_t be an arbitrary total causal mapping over E and let $\Phi_p = \langle V_p, E_p, \phi_p\rangle \equiv \text{COA}(E)$. Then* $\text{DiG}(\phi_t)$ *is consistent with* $\text{PCG}(\Phi_p)$.

Proof. First we show that ϕ_t is consistent with Φ_p. The result follows from Lemma 1.

Satisfaction of condition 13.1: We prove this result by induction. We label the associations in ϕ_p as $\langle \text{Params}(E_j^{(i)}), \hat{E}_j^{(i)}\rangle$ where $E_j^{(i)}$ is the jth minimal self-contained subset found by COA in the ith level of recursion (e.g., the equations for the exogenous variables can be labeled as $E_1^{(0)}, E_2^{(0)}$, etc.). If v is an arbitrary variable such that $v \in \text{Params}(E_j^{(i)})$, we must show that v gets mapped to some equation $e \in \hat{E}_j^{(i)}$. Let $\langle \text{Params}(E_l^{(k)}), \hat{E}_l^{(k)}\rangle$ be an arbitrary association made by COA. Assume that condition 13.1 holds for all associations $\langle \text{Params}(E_j^{(i)}), \hat{E}_j^{(i)}\rangle$ with all $i < k$. We show that it must also hold for the association $\langle \text{Params}(E_l^{(k)}), \hat{E}_l^{(k)}\rangle$. Let $\langle v, e\rangle \in \phi_t$ be an arbitrary association such that $e \in \hat{E}_l^{(k)}$. By definition of a causal mapping, $v \in \text{Params}(e)$ and, therefore, it must be the case that $v \in \text{Params}(\hat{E}_l^{(k)})$. However, according to the induction hypothesis, all $v \in \text{Params}(\hat{E}_l^{(k)}) \setminus \text{Params}(E_l^{(k)})$ have already been assigned to equations; therefore it must be the case that $v \in \text{Params}(E_l^{(k)})$. To complete the induction step, notice that for any association in the initial level of recursion $\langle \text{Params}(E_l^{(0)}), \hat{E}_l^{(0)}\rangle$, it must be the case that $\text{Params}(E_l^{(0)}) \equiv \text{Params}(\hat{E}_l^{(0)})$ so for any $\langle v, e\rangle \in \phi_t$ it must be the case that $v \in \text{Params}(E_l^{(0)})$.

Satisfaction of condition 13.2: Let $\langle v, e\rangle$ be an elementary association in ϕ_t. Consider the association $\langle \text{Clust}(v), \phi_p(\text{Clust}(v))\rangle \in \phi_p$. By condition 13.1, there exists an elementary association $\langle v, e'\rangle \in \phi_t$ such that $e' \in \phi_p(\text{Clust}(v))$. But since ϕ_t is one-to-one, there can be only one equation associated with v. Therefore $e' = e$ (and $\text{Clust}(e') = \text{Clust}(e)$). □

The following corollary shows that if COA returns a total mapping then the causal graph is unique and acyclic:

Corollary 1. *Let E be a well-defined self-contained structure, let $\Phi_p = \langle V_p, E_p, \phi_p\rangle \equiv \text{COA}(E)$, and let ϕ_t be an arbitrary total causal mapping. The following are equivalent*:

(1) *ϕ_p is a total causal mapping.*
(2) $\text{PCG}(\Phi_p)$ *is isomorphic to* $\text{DiG}(\phi_t)$.
(3) $\text{DiG}(\phi_t)$ *is acyclic.*

In this case COA produces necessary and sufficient edges for the system.

4. Conclusions

A standard causal model requires the specification of a set of structural equations and a mapping from equations to variables. COA requires only the former, and produces at least a partial version of the latter. Thus, COA provides a simple, tractable method to extract the causal features of the system that are necessary given the structural equations alone.

Anyone who has attempted to specify the causal structure from a system of more than five or so equations understands that the task quickly gets intractable by hand. This fact is especially true when, as is typically the case, the system is under-constrained. Such a system has several possible causal structures depending on which variables are made exogenous (cf. Fig. 3). For example, a system of N_v variables and $N_e < N_v$ equations has in the worst case N_v-choose-$(N_v - N_e)$ different ways to assign exogenous variables. Without the benefit of COA all of these combinations are viable. This fact restricts the utility of basing causal models on equations and has generally led research to focus on learning causal relations from data. This approach, however, wastes a huge repository of knowledge of physical, social, biological, economic, psychological, etc., relationships that have been discovered by decades of research. The presence of a provably sound causal ordering algorithm provides a practical way to incorporate this wealth of existing knowledge into causal models.

Aside from constructing models, COA provides us a flexible tool for specifying complex interactions with a system. What if we want to swap one part of our system with a new one that depends on different factors? What if we want to manipulate some variables but release others? The standard *Do* operator does not address these complex types of manipulation, but COA handles them seamlessly. COA thus serves as a generating function for arbitrary manipulations on a causal system.

Acknowledgements

We would like to thank Tsai-Ching Lu, Hans van Leijen and Jeroen Boegers for useful discussions on this topic. This research was supported by the National Aeronautics and Space Administration under the Graduate Students Research Program (GSRP), grant number S99-GSRP-085, the Air Force Office of Scientific Research under grants F49620-00-1-0122, F49620-03-1-0122 and FA9550-06-1-0243 to the University of Pittsburgh, by Intel Research and by the National Science Foundation under Faculty Early Career Development (CAREER) Program, grant IRI-9624629.

References

[1] G.F. Cooper, E. Herskovits, A Bayesian method for the induction of probabilistic networks from data, Machine Learning 9 (4) (1992) 309–347.
[2] D. Dash, Caveats for causal reasoning, PhD thesis, Intelligent Systems Program, University of Pittsburgh, Pittsburgh, PA, April 2003.
[3] D. Dash, Restructuring dynamic causal systems in equilibrium, in: R.G. Cowell, Z. Ghahramani (Eds.), Proceedings of the Tenth International Workshop on Artificial Intelligence and Statistics (AIStats 2005), Society for Artificial Intelligence and Statistics, 2005, pp. 81–88.
[4] J. de Kleer, J. Seely Brown, A qualitative physics based on confluences, Artificial Intelligence 24 (1–3) (December 1984) 7–83.
[5] J. de Kleer, B.C. Williams, Reasoning about multiple faults, in: Proceedings of the 5th National Conference on Artificial Intelligence (AAAI-86), Philadelphia, PA, American Association for Artificial Intelligence, 1986, pp. 132–139.
[6] M.J. Druzdzel, H.A. Simon, Causality in Bayesian belief networks, in: Proceedings of the Ninth Annual Conference on Uncertainty in Artificial Intelligence (UAI-93), San Francisco, CA, Morgan Kaufmann Publishers, Inc., 1993, pp. 3–11.
[7] M.J. Druzdzel, H. van Leijen, Causal reversibility in Bayesian networks, Journal of Experimental and Theoretical Artificial Intelligence 13 (1) (January 2001) 45–62.
[8] A.S. Goldberger, Structural equation methods in the social sciences, Econometrica 40 (6) (November 1972) 979–1001.
[9] C.W.J. Granger, Investigating causal relations by econometric models and cross-spectral methods, Econometrica 37 (1969) 424–438.
[10] T. Haavelmo, The statistical implications of a system of simultaneous equations, Econometrica 11 (1) (January 1943) 1–12.
[11] Y. Iwasaki, H.A. Simon, Causality in device behavior, Artificial Intelligence 29 (1) (July 1986) 3–32.
[12] Y. Iwasaki, H.A. Simon, Theories of causal ordering: Reply to de Kleer and Brown, Artificial Intelligence 29 (1) (July 1986) 63–67.
[13] H. Kiiveri, T.P. Speed, Structural analysis of multivariate data: A review, in: S. Leinhardt (Ed.), Sociological Methodology 1982, Jossey-Bass Publishers, San Francisco, 1982, pp. 209–289, Chapter 6.
[14] H. Kiiveri, T.P. Speed, J. Carlin, Recursive causal models, Journal of the Australian Mathematical Society 36 (1984) 30–52.
[15] S.L. Lauritzen, D.J. Spiegelhalter, Local computations with probabilities on graphical structures and their application to expert systems, Journal of the Royal Statistical Society, Series B (Methodological) 50 (2) (1988) 157–224.
[16] T.-C. Lu, M.J. Druzdzel, T.-Y. Leong, Causal mechanism-based model construction, in: Proceedings of the Sixteenth Annual Conference on Uncertainty in Artificial Intelligence (UAI-2000), San Francisco, CA, Morgan Kaufmann Publishers, Inc., 2000, pp. 353–362.
[17] P. Pandurang Nayak, Causal approximations, Artificial Intelligence 70 (1–2) (1994) 1–58.
[18] J. Pearl, T.S. Verma, A theory of inferred causation, in: J.A. Allen, R. Fikes, E. Sandewall (Eds.), KR-91, Principles of Knowledge Representation and Reasoning: Proceedings of the Second International Conference, Cambridge, MA, Morgan Kaufmann Publishers, Inc., San Mateo, CA, 1991, pp. 441–452.
[19] J. Pearl, Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference, Morgan Kaufmann Publishers, Inc., San Mateo, CA, 1988.
[20] J. Pearl, Causal diagrams for empirical research, Biometrika 82 (4) (1995) 669–710.
[21] J. Pearl, Causality: Models, Reasoning, and Inference, Cambridge University Press, Cambridge, UK, 2000.
[22] Q. Shen, T. Peng, R. Milne, Dimensional analysis based causal ordering, in: Proceedings of the 13th International Workshop on Qualitative Reasoning (QR-99), Loch Awe, Scotland, 1999, pp. 193–202.
[23] H.A. Simon, J.R. Kalagnanam, M.J. Druzdzel, Performance budget planning: The case of a research university, Technical report, School of Information Sciences, University of Pittsburgh, Pittsburgh, PA, 2000.
[24] H.A. Simon, Causal ordering and identifiability, in: W.C. Hood, T.C. Koopmans (Eds.), Studies in Econometric Method, in: Cowles Commission for Research in Economics Monograph No. 14, John Wiley & Sons, Inc., New York, NY, 1953, pp. 49–74, Chapter III.
[25] H.A. Simon, Spurious correlation: A causal interpretation, Journal of the American Statistical Association 49 (267) (September 1954) 467–479.
[26] H.A. Simon, Nonmonotonic reasoning and causation: Comment, Cognitive Science 15 (2) (April–June 1991) 293–300.

[27] P. Spirtes, C. Glymour, An algorithm for fast recovery of sparse causal graphs, Social Science Computer Review 9 (1991) 62–72.
[28] P. Spirtes, C. Glymour, R. Scheines, Causation, Prediction, and Search, Springer Verlag, New York, 1993.
[29] P. Spirtes, C. Glymour, R. Scheines, Causation, Prediction, and Search, second ed., The MIT Press, Cambridge, MA, 2000.
[30] R.H. Strotz, H.O.A. Wold, Recursive vs. nonrecursive systems: An attempt at synthesis; Part I of a triptych on causal chain systems, Econometrica 28 (2) (April 1960) 417–427.
[31] J.L. Top, H. Akkermans, Computational and physical causality, in: Proceedings of the Twelfth International Joint Conference on Artificial Intelligence (IJCAI-1991), 1991, pp 1171–1176.
[32] L. Travé-Massuyés, R. Pons, Causal ordering for multiple mode systems, in: Proceedings of the 11th International Workshop on Qualitative Reasoning (QR-97), 1997, pp. 203–214.
[33] N. Wermuth, S.L. Lauritzen, Graphical and recursive models for contingency tables, Biometrika 72 (1983) 537–552.
[34] N. Wermuth, Linear recursive equations, covariance selection and path analysis, Journal of the American Statistical Association 75 (1980) 963–972.
[35] B.C. Williams, Qualitative analysis of MOS circuits, Artificial Intelligence 24 (1984) 281–346.
[36] H. Wold, Causality and econometrics, Econometrica 22 (2) (April 1954) 162–177.
[37] S. Wright, The method of path coefficients, Annals of Mathematical Statistics 5 (1934) 161–215.